DOCUMENT RESUME

ED 283 877 TM 870 416

AUTHOR Jones, Patricia B.; And Others
TITLE Dimensionality Assessment for Dichotomously Scored Items Using Multidimensional Scaling.
PUB DATE Apr 87
NOTE 40p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cluster Analysis; Correlation; Difficulty Level; Error of Measurement; Factor Analysis; *Latent Trait Theory; Mathematical Models; Matrices; *Monte Carlo Methods; *Multidimensional Scaling; Simulation; Test Items
IDENTIFIERS *Dichotomous Variables; Eigen Values

ABSTRACT

In order to determine the effectiveness of multidimensional scaling (MDS) in recovering the dimensionality of a set of dichotomously-scored items, data were simulated in one, two, and three dimensions for a variety of correlations with the underlying latent trait. Similarity matrices were constructed from these data using three margin-sensitive and three margin-free coefficients and used as input to MDS. Stress (S1), S1 by dimension plots, and plots of the scaled items were examined to determine the effect of varying the magnitude and pattern of correlations. The results suggested that items with similar patterns of correlations tend to cluster together, that distance from the center of a cluster is a function of the amount of random error in the item, and that as the number of latent traits underlying the data increases, the dimensionality of the representational space increases. Cluster analysis using MDS coordinates is suggested to isolate homogeneous sets of items, whereas consideration of the S1 coefficient is recommended to determine the number of latent traits in the data. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED283877

Dimensionality Assessment for Dichotomously Scored Items Using Multidimensional Scaling

Patricia B. Jones
Darrell L. Sabers
Michael Trosset

University of Arizona

Paper presented at the Annual Meeting of
the American Educational Research Association
Washington, D. C.
April 1987

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P. Jones

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 870 416

BEST COPY AVAILABLE

ABSTRACT

In order to determine the effectiveness of multidimensional scaling (MDS) in recovering the dimensionality of a set of dichotomously-scored items, data were simulated in one, two, and three dimensions for a variety of correlations with the underlying latent trait. Similarity matrices were constructed from these data using three margin-sensitive and three margin-free coefficients and used as input to MDS. Stress (S_1), S_1 by dimension plots, and plots of the scaled items were examined to determine the effect of varying the magnitude and pattern of correlations. The results suggested that items with similar patterns of correlations tend to cluster together, that distance from the center of a cluster is a function of the amount of random error in the item, and that as the number of latent traits underlying the data increases, the dimensionality of the representational space increases. Cluster analysis using MDS coordinates is suggested to isolate homogeneous sets of items, whereas consideration of the S_1 coefficient is recommended to determine the number of latent traits in the data.

Determining what a set of items really measures remains a problem in educational psychology today just as it was fifty years ago. The problem is that differing methodologies for examining the structure of an item set may yield different results, because their assumptions and techniques differ. This is not necessarily a drawback since this may capture the richness of the data. The purpose of this paper is to examine an alternative approach to determining the structure of a set of items using a combination of multidimensional scaling and cluster analysis.

Investigations into the structure of a set of items frequently involve one or more of three questions: whether the items are pervaded by a single underlying trait; how many traits pervade the data if the items are not unidimensional and which items are related to which traits; or which items are sufficiently similar that they could be said to constitute a unidimensional set of items. Assessing sets of items for unidimensionality or forming unidimensional sets of items has become increasingly important since the development and implementation of item response theory (IRT) because most scaling procedures currently require local independence and unidimensionality. No definitive methodology has yet been advanced for testing for unidimensionality and Lord (1980) has pointed out the need for such a test. To the extent that the assumptions of the IRT models are not met, parameter estimates will be inaccurate.

Multidimensional IRT models have been proposed and/or investigated by a number of researchers including Bock and Aitkin (1981) and McKinley and Reckase (1982a, 1982b, 1983). However, in order to estimate parameters of such models, it is necessary to know how many latent traits underlie the data. Thus, techniques for determining the number of dimensions underlying the data are necessary for multivariate as well as univariate models. Classical approaches to test theory have required assessment of dimensionality to confirm the appropriateness of providing a single test score for a subset of items. Content validity studies have required that the test developer determine whether items which are assumed to share a common trait do, in fact, share the same trait. Thus, while the problems of dimensionality assessment have

received much attention in recent years, the need for techniques to permit determining what a set of items measures is not new and is not solely related to IRT.

Linear Techniques

Classical principal components (PCA) and factor analysis (Gnanadesikan, 1977) have frequently been used to determine the underlying structure of the data. In the PCA model, individual responses are assumed to lie in a space whose axes (not necessarily orthogonal) are formed by the p items. The goal of PCA is to form orthogonal linear combinations of the items such that each linear combination accounts for a maximum amount of the residual variance left unexplained by previous linear combinations. The first principal component is essentially a least-squares regression of a latent trait on the item responses; it accounts for more variance than any other possible combination of items. By successively extracting linear combinations which maximally account for residual variance, orthogonal axes are formed. If the matrix of interitem correlations is less than full rank or if it can be determined that only the first k [$< p$] principal components are needed to account for a sufficient proportion of the variance, a space of smaller dimensionality is considered sufficient to contain the structure represented by the original data. The principal components which constitute the axes of the solution may be interpreted as representing unseen factors underlying the data. If a single principal component appears to account for a sufficient portion of the variance, then the set of items is said to be unidimensional.

In the factor analytical model, each observed variable is assumed to be comprised of a weighted linear combination of latent (i.e., unmeasurable) variables called common factors which account for the common variance among the observed variables as well as a unique latent variable (called the specific factor) which accounts for the variance unexplained by the common factors. In this model, the observed response is regressed on the latent responses. The PCA and factor analysis models are very similar in their approaches to the reduction

of the dimensionality of the original data space and will be regarded as essentially the same here.

Although factor analytical techniques would appear to be ideally suited for determining dimensionality, several problems occur in their use. First, there is no set criterion for determining how many factors or principal components to extract in a given analysis. Secondly, factor analysis requires the specific assumption that the observed variables represent linear combinations of the latent variables. This presents problems for dichotomized variables. If a factor analysis is conducted on a matrix of phi correlation coefficients (which implicitly assume that the distributions underlying the variables are true dichotomies), the model is misspecified since the regression of a binary variable on a continuous variable is not linear (McDonald and Ahlawat, 1974; Mislevy, 1986). Since the range of the values of phi coefficients are affected by the pairwise distributions of the means of the binary variables (i.e., the classical item difficulties), the magnitude of the phi coefficient is affected by the item difficulty as well as the strength of the relationships among the variables. As a result, extraneous factors appear which are related to the difficulties of the items and not to any other source of common variance among the items. This clouds the issue of the true dimensionality of the item set.

One potential solution to the problems listed above is to factor analyze a matrix of tetrachoric correlation coefficients, which assume that bivariate normal distributions underly the pairwise distributions of dichotomized responses. While this solves the problem of nonlinear regressions, it brings about other problems. First, if the assumption of bivariate normality is not met, the tetrachoric correlations will not reflect the true relationships among the variables. Second, tetrachoric correlation coefficients are not directly estimable. Simple approximation formulas may be found in many texts (see, e.g., Lord and Novick, 1968), but these are accurate only in the neighborhood of $r = 0.5$. More complex estimation procedures using Gauss-Hermite quadrature or Newton-Raphson iterations have been suggested by Castellan (1966), Kirk (1973), Divgi (1979), and others. While these

techniques improve the accuracy of the computations, they can become unstable when one or more cell proportions of the pairwise item response table is extremely small. Third, tetrachoric correlation coefficients are inappropriate in cases where guessing occurs since the probability of a correct response is influenced not only by latent ability, but chance factors as well, resulting in spurious guessing factors (Carroll, 1945). While the correlations can be corrected for guessing, other research (Reckase, 1981) has shown that over- or under-correcting yields undesirable results. Other problems include large standard errors (requiring the use of very large samples to achieve stability) and the potential for negative eigenvalues (an undesirable feature when eigenvalues are interpreted as amount of variance accounted for).

Recent developments in factor analysis have included the full-information maximum likelihood (ML) approach (Bock and Aitkin, 1981; Bock, Gibbons, and Muraki, 1985) and the generalized least squares approach of Muthén (1978). These techniques are capable of dealing with dichotomized data and provide a statistic for model fit which is asymptotically distributed as chi-square. However, the statistical test is based on distributional assumptions which may be too restrictive for the variables. Furthermore, for tests of moderate size, very large samples are required to insure the accuracy of the asymptotic approximation. In addition, restrictions are placed on the number of items which may be factor analyzed (according to Mislevy (1986), 25 is an upper limit for the GLS procedure) or the number of factors in the solution (1-3 for tests with 60 items in the ML technique).

From the above discussion, it appears that, although conventional factor analysis and PCA are used to investigate the structure of sets of items, the techniques are not without problems when dichotomized data are analyzed. Other procedures for investigating data structure are necessary.

Multidimensional Scaling

All the techniques described above are based on linear models in which the responses to the items are modeled by a weighted sum of latent variables. In contrast, multidimensional scaling (MDS) models may be described as distance models. The fundamental task of MDS is to recover the underlying structure of a set of data points given the interpoint distances, which are assumed to be represented by a matrix of experimentally determined proximity measures δ_{ij}. In MDS, neither the coordinates of the objects being scaled nor the correct number of dimensions (i.e., the correct number of coordinate axes) is known. Proximities may be measures of similarity or dissimilarity. In the case under consideration here, δ_{ij} is a measure of the similarity of items i and j, for example, a correlation coefficient.

For a representational space in k dimensions, a vector $\underline{x}$ is obtained for each item which gives its (unknown) coordinates in the k-dimensional space. Then the distance of object i from j in the space may be given by the Euclidean distance:

$$d_{ij} = \left[\sum_k |x_{ik} - x_{jk}|^2 \right]^{1/2} . \qquad (1)$$

The Euclidean distance, unlike some other distance measures, is invariant under orthogonal axis translations.

In general, d_{ij} will not equal δ_{ij} because of error in the data or because k is less than the true number of dimensions in the space of the data. In nonmetric MDS, the coordinates of all x_{ik} are obtained subject to a monotonicity constraint, i.e., $\delta_{ij} < \delta_{i'j'}$ implies that $d_{ij} \leq d_{i'j'}$. In other words, the rank orders of the distances d_{ij} are essentially the same as the rank orders of the similarities δ_{ij}, a much less restrictive constraint than an equality constraint and in general, representational spaces obtained in nonmetric MDS will be of lower dimensionality than in metric MDS where the equality constraint holds.

To measure how closely the rank orders are preserved in the k-dimensional MDS configuration, a procedure called monotone least squares or monotone regression (Gnanadesikan, 1977) is carried out on

the d_{ij} to yield <u>disparities</u> $\hat{d}_{ij}$. Kruskal has proposed a quantity which he calls the STRESS, S_1, as a badness-of-fit measure:

$$S_1 = \left| \frac{\Sigma (d_{ij} - \hat{d}_{ij})^2}{\Sigma d_{ij}^2} \right|^{\frac{1}{2}} . \qquad (2)$$

Note that the original δ_{ij} are only indirectly represented in S_1 through their effect on the rank orders of the disparities. Variants of this measure have been proposed by Kruskal (1964), Takane, Young, and deLeeuw (1977), and Guttman (1968).

While S_1 represents the fit of the configuration in the k-dimensional space, little is known about its distribution except that S_1 decreases as k increases. Empirical results led Kruskal to suggest that values of .20, .10, .05, .025, and 0.0 represent poor, fair, good, excellent, and perfect fits, respectively. Kruskal and Wish (1978) have suggested that a value of .15 for k = 1 suggests strongly that the data are unidimensional. Suggestions for choosing a representational space have included a scree-like technique examining plots of S_1 vs. k for elbows and selecting a value for k where a sharp bend occurs. Other possible techniques include comparing S_1 vs. k plots for similarity to simulation results for data generated with a known structure.

<u>Using MDS with Dichotomous Data</u>

Given that MDS is capable of recovering the structure in data under a set of less restrictive conditions than those of factor analysis, it would appear to be a potential solution for the problem of dimensionality analysis with dichotomous data. While factor analysis requires the use of correlation coefficients or covariances, MDS only requires that the similarity measures be ordered. Given the less restrictive nature of the hypotheses in MDS and the problems encountered to date in using factor analysis with dichotomous variables, it is surprising that so little research has been done on the use of MDS in this particular application. While studies have been conducted using MDS on dichotomous data to determine the structure of the data (Korpi and Haertel 1984; Thomas 1984; Koch 1983), the efforts

in these studies have been directed toward comparing results with factor analyses on real data or in interpreting MDS results on real data. The true dimensionalities of the data sets are unknown, the studies cannot definitively establish the validity of MDS in determining dimensionality.

The Problem of Dimensionality Assessment

Determining the number of dimensions underlying the structure of the data is more complex than simply determining which representational space has a small enough value of S_1. For example, a two-dimensional plot may reveal a single curvilinear dimension (Gnanadesikan 1977). Thus, the appropriate dimensionality may differ from that suggested by the S_1 coefficients. Alternative techniques to examining values of S_1 have been proposed. One procedure is to conduct a cluster analysis on the scaled items using an hierarchical clustering algorithm to reveal the data structure. Thus, two distinct clusters in a single dimension on an MDS plot might indicate that two latent variables underlie the data. Kruskal and Wish (1978) also suggested combining closeness information from the MDS plots with proximities information. This can be done by connecting all data points in the plots whose proximities exceed a certain threshold value. This technique is especially useful in revealing curvilinear structure.

While Guttman (1965) recommended that these techniques be used in place of analysis of S_1 values, Kruskal and Wish (1978) have suggested that the researcher use any necessary procedure to search for structure in the data. They point out that the problem of dimensionality assessment goes beyond a simple examination of S_1 values or searching for clusters or regions in the data. They also point out that the problem of determining the correct number of dimensions to represent the data (i.e., the number of axes required for the data set) in an MDS scaling is not equivalent to determining the true dimensionality of the data set.

Comparison with Linear Models

The models for data representation described by PCA and factor analysis are linear models. The MDS models are distance models for which interpretation of the axes may differ. In the linear models, n data points which are p-dimensional vectors of responses to the p items are plotted in space. The data points are the responses of the individuals to the items and the axes are the dimensions described by the items. A subspace in which the axes are linear combinations of the items is obtained. These linear combinations constitute the latent variables of the PCA. In MDS, p items are plotted in a space of some unknown dimensionality and a parsimonious representation of that dimensionality is sought.

The axes of the MDS configuration need not represent latent variables and even if they do, they need not represent the same latent variables as the PCA. Thus, MDS must be regarded as an alternative approach to dimensionality assessment rather than a more general version of PCA.

Monte Carlo Simulation Studies

To date, only two studies have been found which compare the effectiveness of MDS with other techniques for recovering dimensionality in artificially-generated, dichotomized data of known dimensionality and structure. Zwick (1986) successfully used MDS to recover the dimensionality of a set of items with Guttman scaling. She suggested that recovery of a single dimension from such items be used as a criterion before using any procedure for dimensionality assessment. However, she found that the results from an MDS analysis of an actual data set were less clear.

Reckase (1981) simulated 24 data sets consisting of responses of 1000 individuals to 50 items. Of these data sets, 19 were unidimensional, 3 had two factors, and 1 each had three and nine factors. Item difficulties were normally or rectangularly distributed; and guessing parameters were assumed constant for the data set or normally distributed around expected values ranging from 0.0 to 0.75. He compared the effectiveness of PCA and factor analysis (with

orthogonal and oblique solutions), MDS, cluster analysis, and item response theory in recovering the true dimensionality of the data sets. For the PCA and factor analysis investigations he used phi, tetrachoric, and corrected tetrachoric correlation coefficients. For the MDS scalings he used 13 different similarity coefficients (agreement, approval, eta, kappa, koppa, phi, $\bar{x}/\bar{x}_{max}$, tetrachoric (corrected and uncorrected for guessing), Yule's Q, Yule's Y, gamma, tau B, and Lijphart's index). MDS techniques were effective with simulated data, but he felt that they failed to give an indication of the dimensionality of a set of real items.

Several important findings emerged from the Reckase study. First, the type of similarity coefficient used affected the magnitude of the value of S_1. "Margin-free" coefficients (those not affected by item difficulty) such as the tetrachoric correlation and Yule's Q, in general, had larger values for S_1 than "margin-sensitive" coefficients such as phi and agreement. Second, the configuration of the items in the MDS plots was affected by the type of similarity coefficient used. Furthermore, guessing tended to degrade the clarity of the plots because it added random error to the similarity coefficients, affecting their rank orderings. Difficult items were affected more than easy items. Reckase reported results for two-dimensional scalings only, so it was not clear what effect additional dimensions would have on the magnitudes of the S_1 values and the configurations of the items in the chosen representational space. Furthermore, since he was primarily interested in the effects of guessing on the plots, he did not vary the strengths of the item-latent dimension relationship within a given number of latent dimensions.

The failure of the Reckase and Zwick studies to achieve interpretable results for a MDS scaling of actual data when simulation results are so promising is puzzling. With this in mind, this study was undertaken to investigate possible reasons for these failures and to determine the role of the strength of the relationship of the item to the latent dimension in the scaling of items. Since MDS has been used to interpret actual data, additional simulations are needed to determine the validity of this technique for interpreting data.

Methodology

The research questions addressed in this study are, specifically,

1. Are methods of dimensionality assessment using MDS effective in recovering the true dimensionality of sets of items which have been dichotomously scored?

2. Are methods of dimensionality assessment using MDS influenced by the number of dimensions underlying the sets of items, the magnitude of the relationship of the items to the latent trait, the distributions of these magnitudes across the latent traits, or the type of similarity measurement used?

3. Does the type of similarity coefficient used in MDS affect the decision made in assessing dimensionality?

4. How do methods of dimensionality assessment using MDS compare with PCA in terms of recovering the true dimensionality of a set of dichotomously scored items? Do they eliminate the presence of so-called "difficulty" factors?

Data Generation

Data of known dimensionality were obtained by simulating individual responses to a set of hypothetical items under a multidimensional extension of the two-parameter logistic model (2PL) (Birnbaum, 1968):

$$P(X_{ij} = 1 \mid \underline{\theta}_j) = \frac{1}{1 + e^{-1.7\underline{a}_i' (\underline{\theta}_j - \underline{b}_i)}} \qquad (3)$$

This model is essentially the same as the multivariate two-parameter logistic model proposed by Reckase and McKinley (1983) who modified the model somewhat to cope with problems of estimation.

In order to replicate test conditions approximately, item location parameters (b_i) were generated using the standard normal distribution (for data sets with one, two, or three dimensions) or the uniform U(-3,+3) distribution (for one-dimensional data sets). Item discrimination parameters were specified at fixed values or, in one

case, generated from a uniform distribution ranging between .3 and 2.0). Table 1 presents details of the simulated item sets. When ability can be assumed to have a standard normal distribution in the population and there is no guessing, the relation between the item discrimination parameter a_i and the correlation coefficient with the latent trait (the factor loading given in Table 1) is given by

$$a_i = \frac{\tau_{i\theta}}{(1 - \tau_{i\theta}^2)^{\frac{1}{2}}} \qquad (4)$$

(see Lord, 1980). $\tau_{i\theta}$ represents the correlation of the item with the latent trait.

A vector of abilities (with length determined by the specified dimensionality of the item set) was generated for each individual. Abilities were assumed to be normally and independently distributed across and within individuals, that is, not only was the ability of one subject unrelated to that of another subject, but also ability on one dimension was assumed to have no relationship to ability on any other dimension.

Dichotomized responses to the items were simulated for each of 1000 individuals using (3) and comparing $P(X_{ij} = 1 \mid \theta_j)$ $(= P_{ij})$ to a $U(0,1)$ random variable R_{ij}. If $P_{ij} < R_{ij}$, $X_{ij} = 1$, otherwise $X_{ij} = 0$. Similarity matrices were then generated from cross-classification tables for responses to pairs of items. Six coefficients were selected for this study, three margin-sensitive (agreement, phi, and kappa), and three margin-free ($\bar{s}/\bar{s}_{max}$, Yule's Q, and the tetrachoric correlation). It was felt that these six coefficients would give an adequate representation of the types of results to be expected. Their formulas are given in Reckase (1981).

Each similarity matrix was scaled using nonmetric MDS analyses in one to five dimensions. Euclidean distances were specified for comparability with other analyses and because computational algorithms are more efficient and more robust when the Euclidean metric is used. The scaling was carried out using the Kruskal algorithm in the MDS

procedure in the SYSTAT statistical package (Wilkinson 1984) which is available for the IBM PC.

For comparison with results which would be obtained using traditional methodologies for dimensionality assessment, PCA was carried out using the FACTOR package in Systat. (This package does not include a traditional factor analysis package, but does permit rotation of a reduced set of principal components.) Analyses were carried out on both phi and tetrachoric correlation matrices. The magnitudes of the eigenvalues obtained in the analysis were examined to determine the appropriate number of principal components to retain.

Values of S_1 and the S_1 by k plots were examined to determine whether they suggested an appropriate size for the representational space in MDS. Two- and higher-dimensional plots of scaled results were examined for configurations suggesting the dimensionality of the data.

Results of Analyses

Choosing a Representational Space

The magnitude and the behavior of the S_1 coefficient were found to be a function of the number of dimensions in the data, the amount of random error in the data, the type of similarity coefficient used, and, for margin-sensitive coefficients, the distribution of item difficulties. S_1 increased as the amount of random error in the data increased (i.e., as the sum of the squared factor loadings decreased). Furthermore, as the amount of error increased, the change in slope in the graphs of S_1 vs. the number of dimensions also decreased. This made determining the appropriate number of dimensions more difficult. Figure 1 presents the S_1 values obtained by scaling data sets 19-21 in one through five dimensions using an input matrix of agreement coefficients. These data sets are three-dimensional with similar factor structure; however, the sum of the squared factor loadings are .81, .25 and .09, respectively. The change in slope of the graph is abrupt at a three-dimensional solution for data set 19, less abrupt for data set 20, and difficult to determine for data set 21. The magnitude

of S_1 rises steadily across these data sets for a given dimensionality. Figure 2 represents S_1 values obtained by scaling matrices of tetrachoric coefficients for the same data set. A clear difference appears between the graphs for data sets 19 and 20 and that for data set 21. This graph was somewhat unusual for two reasons. The magnitude of S_1 increased slightly with increasing dimensionality in the representational space for data set 19. This effect was encountered occasionally in situations where the magnitude of S_1 was low and was considered of no importance. Secondly, the magnitude for S_1 was approximately the same for data set 19 as for data set 20. The one clear difference is the lack of an elbow at a two-dimensional solution for data set 20. Conclusions regarding the number of dimensions in the representational space be difficult to make for data set 21. Furthermore, the number of dimensions would differ according to whether the agreement or the tetrachoric coefficient was being used.

For unidimensional data sets 1-9 a one-dimensional representational space gave low values of S_1 [< .15] when margin-sensitive coefficients were scaled. As the amount of random error in the data increased, S_1 increased. For the margin-free coefficients, S_1 was very high (.4-.5) for a one-dimensional solution. It tended to drop off steadily with no sign of an elbow. This precludes the use of a criterion value such as .15 for determining the appropriate number of dimensions in the representational space when margin-free coefficients are used. However, such a criterion might be useful if margin-sensitive coefficients were scaled.

For the two-dimensional data, an examination of S_1 for margin-sensitive coefficients suggested a two-dimensional representational space. However, it was more difficult to determine the appropriate dimensionality for data sets with substantial random error (e.g., 12). In these data sets, S_1 values were high and no elbow was clear in the graph. Margin-free coefficients generally had low values of S_1 for a one-dimensional representational space, although the value, again, was affected by the amount of random error. In data sets 15-18, the S_1 coefficient behaved as if the data were unidimensional.

The three-dimensional data yielded solutions which were similar for the two-dimensional data except that one additional dimension was necessary in the solution. This suggests that when a data set has k latent traits, k dimensions are required for scaling margin-sensitive coefficients and (k-1) dimensions are required for margin-free coefficients.

Of the margin-sensitive similarity coefficients used, the agreement coefficient had the lowest S_1 values and was the least sensitive to random error. When coordinates were examined for items scaled using a matrix of agreement coefficients, it was clear that one dimension in the representational space was simply item difficulty. While the kappa and phi coefficients were also sensitive to item difficulty, the effect was less strong, since the effect of item difficulty on these coefficients is a restriction of range. While it was not difficult in most cases to determine an appropriate number of dimensions in the representational space using the agreement coefficient, it was considerably more difficult using the kappa and phi coefficients. For unidimensional data, agreement coefficients ranged between .02 (for data set 4) and .29 (data set 9).

Data sets with rectangularly distributed item difficulties tended to have lower S_1 values than their normally distributed counterparts. This is probably because the items are uniformly distributed along the scale of difficulty. As a result, they are more spread out, making reversals in rank ordering less likely.

Although the tetrachoric correlation coefficient tended to yield S_1 values which were similar to those for Yule's Q, it occasionally yielded much smaller values when factor loadings were high. Examination of scale values and item plots indicated that a difficulty dimension was present for the tetrachoric correlation. This was not an artifact of the scaling procedure, since the same effect was noted when PCA was carried out on the matrix of tetrachoric correlations.

No simple recommendations can be made regarding choice of an appropriate representational space from these results. When a sharp elbow is present in the plot of S_1 by dimensions, a choice of the appropriate number of dimensions is clear. If the initial level of

stress is low, a one-dimensional representational space is appropriate. However, as will be seen from an examination of item plots, the interpretation of this space differs somewhat between margin-sensitive and margin-free coefficients. When no low value of stress is reached for a reasonable (as defined by the researcher) number of dimensions, item plots will have to be examined to determine whether they cast additional light on the issue.

Results of Examination of Eigenvalues

PCA tended to yield good results with the simulated data. This was expected, since the data were generated using a model for which PCA should be maximally effective. This might not be the case for data with a nonlinear underlying model.

The magnitudes of the eigenvalues were affected by the number of dimensions in the data, the amount of random error present in the data, the difficulty level of the items, and, when an item loaded on multiple factors, the discrepancies between those loadings. Except in the cases of data sets containing fewer factor structures than factors (for example, data sets 16-18), PCA yielded solutions with as many or more large eigenvalues than factors. Spurious factors due to difficulty were noted for analyses of both phi and tetrachoric correlation coefficients; these were more likely to occur when correlations with the latent factor were high. The tetrachoric matrices in such instances tended to be non-Gramian and the eigenvalues representing systematic effects were substantially higher than those obtained in analyzing the corresponding matrix of phi coefficients. Figure 3 presents graphs of eigenvalues obtained in PCA of three-dimensional data sets 19-21 using phi coefficients. Figure 4 presents results of analyses on the same data sets using tetrachoric correlations. As the amount of random error increased in the data, the magnitudes of the eigenvalues of the first three principal components decreased. Since the phi coefficients tended to yield lower eigenvalues for the primary principal components than the tetrachoric coefficients, result, the elbows in the graph appeared less well-defined.

Two additional things were found to affect the magnitude of eigenvalues corresponding to principal components regardless of whether phi or tetrachoric correlations were analyzed. First of all, the magnitudes are affected by the heterogeneity of the correlations with the latent trait in multidimensional sets. In data sets 10-12 and 19-21, each item loads on only one of the k (= 2 or 3) factors in the item set and the first k principal components are approximately equal. In data sets 13-14 and 22-25, as the k factor loadings become more homogeneous, a single principal component tends to dominate the eigenvalues. Secondly, the magnitudes of the principal components are affected by the heterogeneity of the items. Data sets 16-18 are very similar to data sets 13-15 except that all items have the same factor loadings on each factor. Whereas data sets 13-15 had a dominant factor and a smaller second factor, data sets 16-18 had only one dominant factor. These findings have two implications. First, the principal components need not correspond to the original generating latent traits. This has implications for interpretation of principal components. Second, the dimensionality of an item is a function of the set of items within which it occurs. This is an example of Bejar's (1983) statement that dimensionality, like reliability, is situation specific. Clearly, this has important implications for test equating.

While data sets 13-15 appeared to have a single dominant factor and a substantially smaller second factor (according to eigenvalue analysis), many researchers would declare the set of items to be unidimensional. On the other hand, S_1 by dimension plots clearly showed that these items were behaving in a similar fashion to data sets 10-12. It may be, then, that MDS, because it does not attempt to form linear combinations, is better at recognizing dimensionality. S_1 by dimension plots for data sets 16-18, however, behaved like those for unidimensional data sets. The reason for this will become clear in the next section.

Examination of MDS Plots

Items are located in the MDS representational space according to their similarity. The items with the highest similarity coefficients

will be the items with the most similar factor loadings. As a result, items with similar factor loadings tend to cluster together. This effect was apparent in all data sets regardless of the type of similarity coefficient used. However, the margin-sensitive coefficients contained an extra dimension due to difficulty which pulled the clusters apart giving a linear appearance. The appearance of the line varied with the type of margin-sensitive coefficient input into the scaling algorithm. Since the scaling results are essentially the same except for differences in appearance caused by the difficulty dimension, results will be discussed for the margin-free coefficients first. Data sets 1-9 all yielded a single cluster or disk of items. As the amount of random error increased, the disk tended to pull apart, but it was still recognizable as a single structure. Data sets 7-9 contained items with different factor loadings on the single factor. Figure 5 presents results obtained for data set 8 for a scaling of Yule's Q coefficients. Items labeled 1, 2, and 3 correlated .9, .6, and .3 with the latent trait, respectively. When the items in these data sets were scaled, the items most closely related to the latent trait appeared on the interior of the cluster in a small region. As the relationship of the items to the latent trait weakened, the items were moved away from the center of the cluster and spaced in concentric circles about the center. This suggests that the magnitude of the factor loading controls the tightness of the clustering.

As the number of dimensions in the item set increased, the number of clusters increased. Data sets 10-15 yielded two clusters of items; data sets 19-24 yielded three clusters of items. Figure 6 shows scaling results obtained for a matrix of tetrachoric coefficients for data set 19 which had factor loadings of .9 on one of three factors. Items related to the first, second, or third factor are identified with a 1, 2, or 3. Three clusters are clearly visible here (although two items trail from the clusters).

As noted previously with the unidimensional data, the tightness of the clusters decreased as the amount of random error in the data increased. Examination of the item plots suggested why the one-dimensional MDS solutions for the margin-free coefficients had such

high S_1 values. When two- and higher-dimensional data are scaled, differences among items due to random error are small compared to differences due to association with different combinations of factors. As a result, rank orderings of item distances are likely to contain fewer large differences in rank orderings resulting in a lower S_1 value. When one-dimensional data are scaled, differences among items are caused only by random error. Rank orderings are random and result in a high S_1 value. For n objects, n-1 dimensions are required to account for the differences in a satisfactory manner.

When the correlations are low, there are substantial amounts of random error to account for, and, accordingly, these items are scaled farther from the cluster center. For the multidimensional data sets, the differentiation between the clusters tended to decrease as the amount of random error increased. This is because the systematic error component (the component which influences the separation of the cluster centers) becomes small relative to the random error component (the component which influences the distance of the items from the cluster centers). As a result, the less reliable items tend to overlap.

When the three-dimensional data were scaled, clusters were located such that items with identical loadings (but on different factors, as in data sets 19-21) were equally distant from each other. Whereas this can be done on a line for two-factor data, an equilateral triangle is required for three-factor data and the representational space will be required to have two dimensions. This suggests that a k-dimensional latent space will require (k-1) dimensions in an MDS space (unless a difficulty dimension is required). The vertices of the triangle may be considered to represent the simplest factors. For example, items 1-13 in data set 19 would form a cluster at one vertex, items 14-27 would form another cluster at a second vertex, and the remaining items would form the third cluster at the last vertex. The number of distinct factor structures in the data controls the number of clusters. For this reason, data sets 13-15 will form two clusters even though the items are highly related to each other. However, their separation will be controlled by the relative strengths of their loadings on the latent

variables; the differentiation of the clusters will be controlled by the magnitudes of the loadings on the latent variables.

Data sets 27-29 provide an illustration of the above point. Two-dimensional MDS plots are presented for these data in Figures 7-9. In data set 27, three clusters are distinctly visible. Items which are related to the first factor (labeled 1) are located in a cluster on the left of the diagram; items related to the second factor cluster (labeled 3) on the right. Items which are related to both factors (labeled 2) are located between the two single-factor clusters. Since they are equally related to both factors, they are spaced midway along the line joining the centers of the two single-factor clusters. Items 1-13 and 28-40 are highly related to their respective factors and their dispersal from the cluster center is small. The two latent factors account for only 50% of the systematic variance in items 14-27 (as opposed to 81% in the remaining items), so the cluster appears more diffuse. In data set 28, items 1-13 (labeled 1) were more heavily related to the first factor than the second (.7, .3), items 14-28 (labeled 2) were related equally to both factors (.5, .5) and items 28-40 (labeled 3) were more heavily related to the second factor than the first (.3, .7). The results in Figure 8 show that, again, three clusters were formed with the items equally related to both factors located toward the center of the plot. However, the clusters on the left and right sides of the plot were more spread out than in data set 27 where factor loadings accounted for 81% of the variability. In data set 28, the factors only account for 58% of the variability in the first and third clusters. Differentiation among the clusters was accordingly less clear.

Data set 29 yielded the most informative results. The plot obtained by scaling tetrachoric correlations for this data set is shown in Figure 9. Here there are seven distinct factor structures, with items loading on one, two, or three factors. Single factor items were scaled in clusters according to their factor relationship (i.e., items 1-6, 7-12, and 13-18 were scaled in three separate clusters labeled 1, 2, or 3). These clusters were equally spaced as in data sets 19-21. The clusters containing homogeneous sets of items which loaded on two

of the three factors (labeled 4, 5, or 6) were located near a line drawn between the two vertices of the triangle containing the factors to which the items were related. Thus, if a cluster contained items related to factors 1 and 2, it was located between the cluster for factor 1 and that for factor 2. The cluster was located closest to the cluster to which it was most highly related. Therefore, a cluster containing items generated from (.7, .3, .0) factor loadings was located between the clusters generated by (.9, .0, .0) and (.0, .9, .0) factor loadings, but closer to the former than the latter. The cluster containing items generated by (.5, .5, .5) factor loadings (labeled 7) was located at the interior of the triangle. As items take on a more complex factor structure, they will be pulled away from a line or a point toward other factors on which they load. This suggests that a tetrahedron in three dimensions will be required to represent a four-factor structure and that items will be located in the interior of the tetrahedron if they relate to all four factors.

These findings have important implications for assessment of dimensionality in MDS. An examination of the clustering in MDS plots will help to isolate sets of homogeneous items. However, if the goal of the MDS procedure is to determine the number of latent dimensions present in the data, the magnitude and behavior of the S_1 coefficient must also be considered.

Plots for the margin-sensitive coefficients appeared quite different because of the presence of a dimension related to difficulty. When unidimensional item sets were scaled using agreement, kappa, or phi coefficients, items were arranged in a line with the most difficult items at one end and the least difficult at the other. For the agreement coefficient, the line was straight. However, for the kappa and phi coefficients, the line was in the shape of a horseshoe. An example of the type of plot obtained is shown in Figure 10, which shows the results of scaling kappa coefficients in data set 1. Items are ordered by difficulty with A the easiest item, progressing through the alphabet to numbers and symbols. Note that while two dimensions are required to represent the graph satisfactorily, a single curvilinear dimension could be hypothesized for the data. The kappa coefficient

was specifically suggested to deal with the problems of chance agreement due to marginal distributions, so it should be less sensitive to item difficulty than the agreement coefficient. However, both kappa and phi have ranges which are restricted by differences in item difficulties; hence, their values are still affected by item difficulty. This is especially true at the extremes of the distribution and may help to account for the curvilinear effect.

As the correlation of each item with the latent trait decreased, the clarity of the plots tended to decrease. The phi coefficient was somewhat more affected by this problem than kappa. As the random error increased, items tended to vary from the line or horseshoe degrading the clarity of the original shape. Plots using kappa coefficients tended to take on curvilinear forms other than a horseshoe, while plots using agreement coefficients tended to resemble a filled-in ellipse. This increase in scatter about the "true" item location has already been described for cluster-type plots.

When data contained two or more dimensions, plots using agreement coefficients took on a circular or elliptical shape. Figure 11 presents results for scaling a matrix of agreement coefficients for data set 19. Items are labeled according to the factor on which they load most highly, as can be seen, items are scaled in arcs about an axis. Examination of the item difficulties showed that the axis was related to item difficulty with easy items scaled at the one end and difficult items at the other. Items which were of extreme difficulty or easiness were closer together than items of moderate difficulty. It appears that such items are more like each other because of their difficulty or easiness than because of any relationship to a factor. However, since an item related to one factor must be as distant from an item of the same difficulty related to another factor as it is from any other item of the same difficulty related to a third factor additional dimensions will be needed to represent the distances in a satisfactory manner. For three-dimensional item sets this structure was similar to a football.

Plots for kappa and phi tended to resemble the plots for the margin-free coefficients as more dimensions were added to the data with

items ordered within the clusters according to difficulty. In general, kappa, phi, and $\bar{E}/\bar{E}_{max}$ tended to have plots which were most influenced by random error in the data and to yield the most variable results. For this reason, their use is not suggested. Results for Yule's Q and the tetrachoric coefficients were very similar and either of the two would yield good results. Except for occasional problems in values for the tetrachoric correlation with a relationship to item difficulty when the correlation with the latent trait was very high, either could be used interchangeably. As previously noted, the agreement coefficient is suggested when examining data for unidimensionality, since only the margin-sensitive coefficients were capable of indicating a one-dimensional solution for such data.

Cluster Analyses

The clusters present in the MDS plots suggested that cluster analysis might be appropriate for classifying items into homogeneous sets. While Reckase (1981) did investigate cluster analysis as a technique for identifying homogeneous item sets, he applied it directly to the same similarity matrices used in the MDS. He found that many items were misclassified even when the correct number of dimensions (or clusters) was used. Furthermore, he found that margin-sensitive coefficients tended to group items of similar difficulty.

Although cluster analysis has a number of problems as a classification technique, two are especially difficult for the practitioner - susceptibility to local solutions (in some cases, due to the presence of outliers) and lack of measures for determining the optimal number of clusters in the solution. However, since the results were clearly clustered, the technique was applied to the coordinates of the items in the MDS representational space (rather than the similarity matrices based on item data). It was hoped that the technique would permit identification of homogeneous item groupings even if it did not give information about the number of dimensions in the data. In addition, using the MDS coordinates might reduce the sensitivity to outliers. The first two cluster analysis procedures used were single linkage and complete linkage options in the CLUSTER procedure in the

SPSS-X statistical package (SPSS-X User's Guide, 1986). Euclidean distance measures were used with the data. It was hoped that one of these procedures would be successful with the MDS results obtained from the agreement matrices. However, neither of these techniques proved useful in classifying items into clusters. Classification appeared to be random and these techniques were dropped. The scaling coordinates obtained for matrices of tetrachoric coefficients for data sets 27-29 (two- and five-dimensional solutions) were input to the k-means clustering program BMDPKM (Dixon, 1985). This technique is more effective with the spherical-type clusters obtained in the MDS analyses and gives information on the within-cluster and between-cluster mean squares for each variable (here, MDS dimensions) to permit assessment of the relative importance of each variable in classifying the data.

In general, results of the k-means clustering were good. No cases were misclassified when three clusters were specified for either two- or five-coordinate MDS solutions in data set 27. When more clusters were specified than were actually needed, the algorithm simply split the larger center cluster into smaller ones. When fewer clusters were specified than the correct number, the center cluster was split so that half the items which loaded equally on both factors were assigned to each of the single-factor clusters. Significance tests on the importance of each variable to classification revealed that only the first dimension was consistently useful in classifying items into groups. As the number of clusters requested increased, additional dimensions were required to separate the items into homogeneous groups, but they were of substantially less importance to the classification procedure.

More problems were encountered with cluster analyses in data set 28 which had less differentiation between the clusters. Initial analyses of both two- and five-dimensional data misclassified about one-third of the cases, primarily because clusters 2 and 3 were split on a horizontal rather than vertical axis. However, when seed values were specified which appeared to be near the cluster centers, no cases were misclassified for the two-dimensional data and only one for the five-dimensional data. A similar problem was encountered for the

seven-cluster solution for data set 29. Although all cases were classified correctly when the two-dimensional scaling results were input, the five-dimensional data gave poor results. However, after seed values were specified, only two cases were misclassified. This suggests strongly that the researcher should try multiple solutions with different seed values, particularly in regions where clusters are suspected. If results are stable for a number of solutions and if the solutions seem reasonable, then the cluster solution may be accepted. These results also suggest that more dimensions than necessary for the MDS representational space should not be input into the clustering procedure.

One additional problem encountered in the analyses was the lack of a criterion for determining the optimal number of clusters. Since the plotting results clearly suggested the correct number of clusters, the problem was not difficult here. The C-index (Seber 1984) yielded poor results and failed to indicate the correct number of clusters. When the statistic was modified to include only those variables which were significant in the clustering, the results were very good. More research needs to be undertaken to determine what, if any, indices of clustering may be useful.

Conclusions

As in the Reckase (1981) study, MDS has been demonstrated to be a useful technique in classifying items into homogeneous groups when the sources of variability in the data are well-specified. While Reckase was interested in the effect of guessing on MDS results, this study has focused on the effect of the relationship of the item with the latent variable. Although some differences have been found which may be due to differences in scaling algorithms, results are in substantial agreement. Reckase found that guessing tended to degrade the quality of the item plots. In this study, it was observed that as item correlations with the latent trait decreased, items were more widely dispersed about the cluster center. In both cases, the dispersion may be attributed to increased random error in the data; for the Reckase study, difficult items were most affected, whereas in the present

study, all items were affected regardless of their difficulty level. Although Reckase had one data set which contained more factor structures than factors (and which yielded as many clusters as structures), for the most part, the two variables were confounded. This study shows clearly that the number of clusters is related to the number of distinct factor structures in the data. The results of this study suggest that the number of factors in the data is related to the number of dimensions in the representational space, but that factors and dimensions must be distinguished in MDS.

Of the coefficients used, three (kappa, phi, and $\bar{s}/\bar{s}_{max}$) were found to be particularly sensitive to random error and their use is not suggested. Of those remaining, the agreement coefficient was found to be useful in cases where unidimensionality is expected in the data. While it is sensitive to item difficulty, this dimension was clear on the item plots. The tetrachoric coefficient and Yule's Q are particularly useful when the goal of the analysis is to isolate homogeneous sets of items since they tended to yield clustered results.

Because of the tendency of items to cluster together according to their similarity, it is suggested that cluster analysis of the MDS coordinates be conducted to isolate homogeneous sets of items. While the tendency of the technique to yield solutions which are sensitive to local problems should not be overlooked, some of the problems may be rectified by using seed values for cluster centers. This has the advantage of permitting the researcher some control over the solution. Because misclassifications tended to occur when MDS representational spaces had too many dimensions, it is suggested that no more than the necessary number of dimensions be used for the analyses. If two-dimensional MDS results are adequate to determine the structure of the data, then no more than two dimensions should be used.

These results deal only with simulated data. Further research should be carried out to determine whether these results are appropriate for actual test data. In addition, research needs to be carried out to determine whether any of the measures of clustering is useful in determining the optimal number of clusters in this situation.

References

Bejar, I. I. (1983). Introduction to item response models and their assumptions. In R. K. Hambleton, ed., Applications of Item Response Theory, Vancouver, BC: Educational Research Institute of British Columbia.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, Statistical Theories of Mental Test Scores, Reading, MA: Addison-Wesley.

Bock, R.D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. Psychometrika, 46, 443-459.

Bock, R. D., Gibbons, R. D. and Muraki, E. (1985). Full information factor analysis (MRC Report No. 85-1). Chicago, IL: National Opinion Research Center.

Carroll, J. B. (1945). The effect of difficulty and chance success on correlations between items or between tests. Psychometrika, 10, 1-19.

Castellan, Jr., N. J. (1966). On the estimation of the tetrachoric correlation coefficient. Psychometrika, 31, 67-73.

Divgi, D. R. (1979). Calculation of the tetrachoric correlation coefficient. Psychometrika, 44, 169-172.

Dixon, W. J. (1985). BMDP Statistical Software. Berkeley, CA: University of California Press.

Gnanadesikan, R. (1977). Methods for statistical data analysis of multivariate observations. New York: John Wiley and Sons.

Guttman, L. (1965). The structure of interrelations among intelligence tests. In Proceedings of the 1964 Invitational Conference on Testing Problems. Princeton, N.J., Educational Testing Services, 25-36.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika, 33, 469-506.

Kirk, D. B. (1973). On the numeric approximation of the bivariate normal (tetrachoric) correlation coefficient. Psychometrika, 38, 259-268.

Koch, W. (1983). The analysis of dichotomous test data using nonmetric multidimensional scaling. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada. (ERIC Document Reproduction Service No. ED 235 204).

Korpi, M. and Haertel, E. (1984). Locating reading comprehension test items in multidimensional space: An alternative analysis of test structure. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 246 108).

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: a numerical method. Psychometrika, 29, 115-129.

Kruskal, J. B. and Wish, M. (1978). Multidimensional Scaling. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-011.

Lord, F. M. (1980). Applications of Item Response Theory to Practical Testing Problems. Hilladale, NJ: Lawrence Erlbaum Associates.

Lord, F. M. and Novick, M. R. (1968). Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley.

McDonald, R. P. and Ahlawat, K. S. (1974). Difficulty factors in binary data. British Journal of Mathematical and Statistical Psychology, 27, 82-99.

McKinley, R. L. and Reckase, M. D. (1982a). MAXLOG: A computer program for the estimation of the parameters of a multidimensional logistic model. Behavior Research Methods and Instrumentation, 15, 389-390.

McKinley, R. L. and Reckase, M. D. (1982b). Multidimensional latent trait models. Paper presented at the National Council on Measurement in Education, New York. (Eric Document Reproduction Service No. ED 220 506).

McKinley, R. L. and Reckase, M. D. (1983). The analysis of dichotomous test data using multidimensional latent trait theory. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. Journal of Educational Statistics, 11, 3-32.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. Psychometrika, 43, 551-560.

Reckase, M. D. (1981). The formation of homogeneous item sets when guessing is a factor in item responses. Office of Naval Research Report 81-5.

Seber, G. A. F. (1984). Multivariate Observations. New York: John Wiley and Sons.

SPSS (1986). SPSS-X User's Guide. Chicago, IL: SPSS, Inc.

Thomas, J. A. (1984). A comparison of factor analysis and multidimensional scaling in determining underlying structure. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 244 994).

Wilkinson, L. (1984). SYSTAT: the System for Statistics. Evanston, IL: Systat, Inc.

Zwick, R. (1986). Assessing the dimensionality of dichotomous item responses: Theoretical and empirical perspectives. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Table 1. Characteristics of One-, Two-, and Three-Dimensional Data Sets.

Data Set Number	Factor Loading 1	2	3	Item Difficulty Distribution	Number of Items
1	.9	-	-	Normal	40
2	.6	-	-	Normal	40
3	.3	-	-	Normal	40
4	.9	-	-	Rectangular	40
5	.6	-	-	Rectangular	40
6	.3	-	-	Rectangular	40
7	.9	-	-	Normal	13
	.6	-	-		14
	.3	-	-		13
8	.9	-	-	Rectangular	13
	.6	-	-		14
	.3	-	-		13
9	U*	-	-	Normal	40
10	.9	.0	-	Normal	20
	.0	.9	-		20
11	.5	.0	-	Normal	20
	.0	.5	-		20
12	.3	.0	-	Normal	20
	.0	.3	-		20
13	.9	.2	-	Normal	20
	.2	.9	-		20
14	.7	.3	-	Normal	20
	.3	.7	-		20
15	.6	.4	-	Normal	20
	.4	.6	-		20
16	.8	.2	-	Normal	40
17	.6	.3	-	Normal	40

* uniformly distributed factor loadings (.3 through .9)

Table 1. (continued)

Data Set Number	Factor Loading 1	2	3	Item Difficulty Distribution	Number of Items
18	.6	.4	-	Normal	40
19	.9	.0	.0	Normal	13
	.0	.9	.0		14
	.0	.0	.9		13
20	.5	.0	.0	Normal	13
	.0	.5	.0		14
	.0	.0	.5		13
21	.3	.0	.0	Normal	13
	.0	.3	.0		14
	.0	.0	.3		13
22	.9	.2	.2	Normal	13
	.2	.9	.2		14
	.2	.2	.9		13
23	.8	.2	.0	Normal	13
	.0	.8	.2		14
	.2	.0	.8		13
24	.5	.3	.1	Normal	13
	.1	.5	.3		14
	.3	.1	.5		13
25	.5	.3	.1	Normal	13
	.4	.2	.0		14
	.3	.0	.1		13
26	.5	.5	.2	Normal	20
	.5	.2	.5		20
27	.9	.0	-	Normal	13
	.5	.5	-		14
	.0	.9			13
28	.7	.3	-	Normal	13
	.5	.5	-		14
	.3	.7	-		13
29	.9	.0	.0	Normal	6
	.0	.9	.0		6
	.0	.0	.9		6
	.7	.5	.0		6
	.0	.7	.5		6
	.5	.0	.7		6
	.5	.5	.5		4

Captions for Figures

Figure 1. S_1 by dimension plot for three-dimensional data sets 19-21 (agreement coefficient).

Figure 2. S_1 by dimension plot for three-dimensional data sets 19-21 (tetrachoric coefficient).

Figure 3. Eigenvalue magnitude for principal components analysis of data sets 19-21 (phi coefficients).

Figure 4. Eigenvalue magnitude for principal components analysis of data sets 19-21 (tetrachoric coefficients).

Figure 5. MDS results for one-dimensional data set (8). 13 items .9, 14 items .6, 13 items .3 (Yule's Q).

Figure 6. MDS Results for three-dimensional data set (19). 13 items (.9, 0, 0), 14 items (0, .9, 0), 13 items (0, 0, .9) (tetrachoric coefficient).

Figure 7. MDS Results for two-dimensional data set (27). 13 items (.9, 0), 14 items (.5, .5), 13 items (0, .9) (tetrachoric coefficient).

Figure 8. MDS Results for two-dimensional data set (28). 13 items (.7, .3), 14 items (.5, .5), 13 items (.3, .7) (tetrachoric coefficient).

Figure 9. MDS results for three-dimensional data set (29). Mixed factor loadings (tetrachoric coefficient).

Figure 10. MDS results for one-dimensional data set, normally distributed item difficulty (kappa coefficient).

Figure 11. MDS results for three-dimensional data set (19) (agreement coefficient).

BEST COPY AVAILABLE

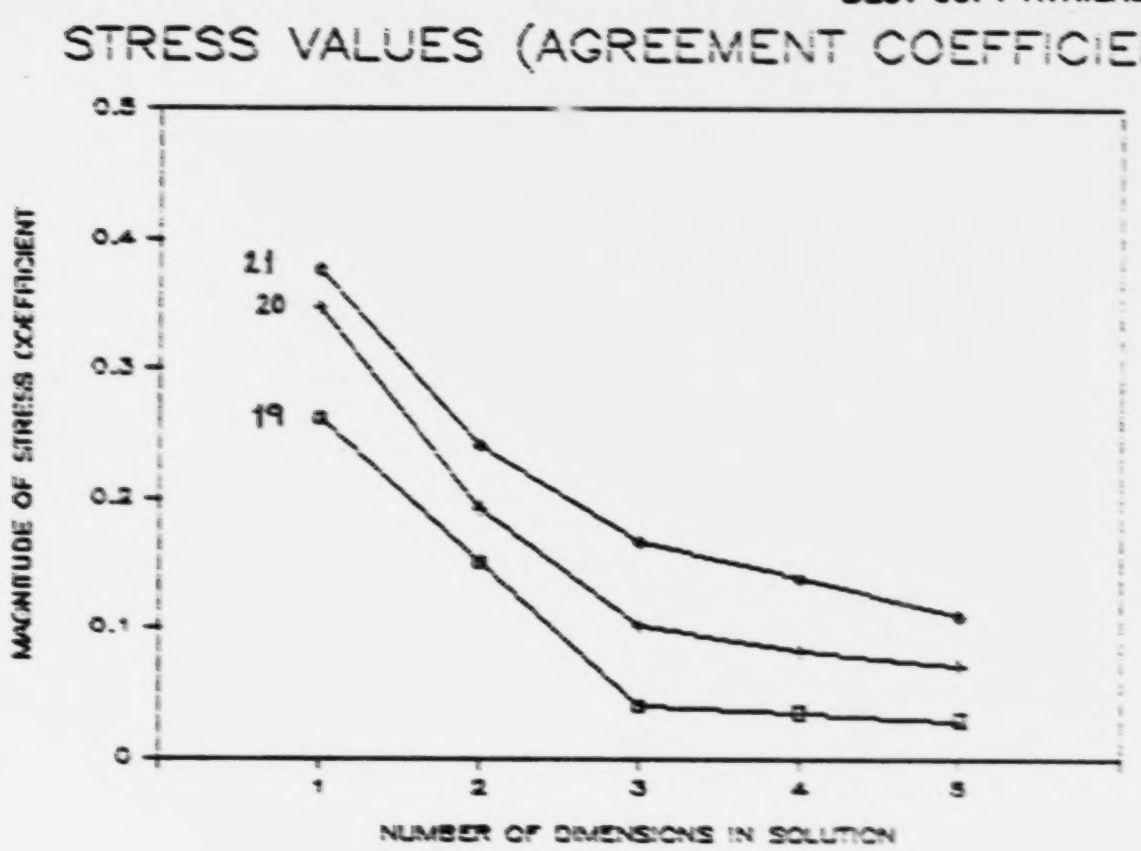
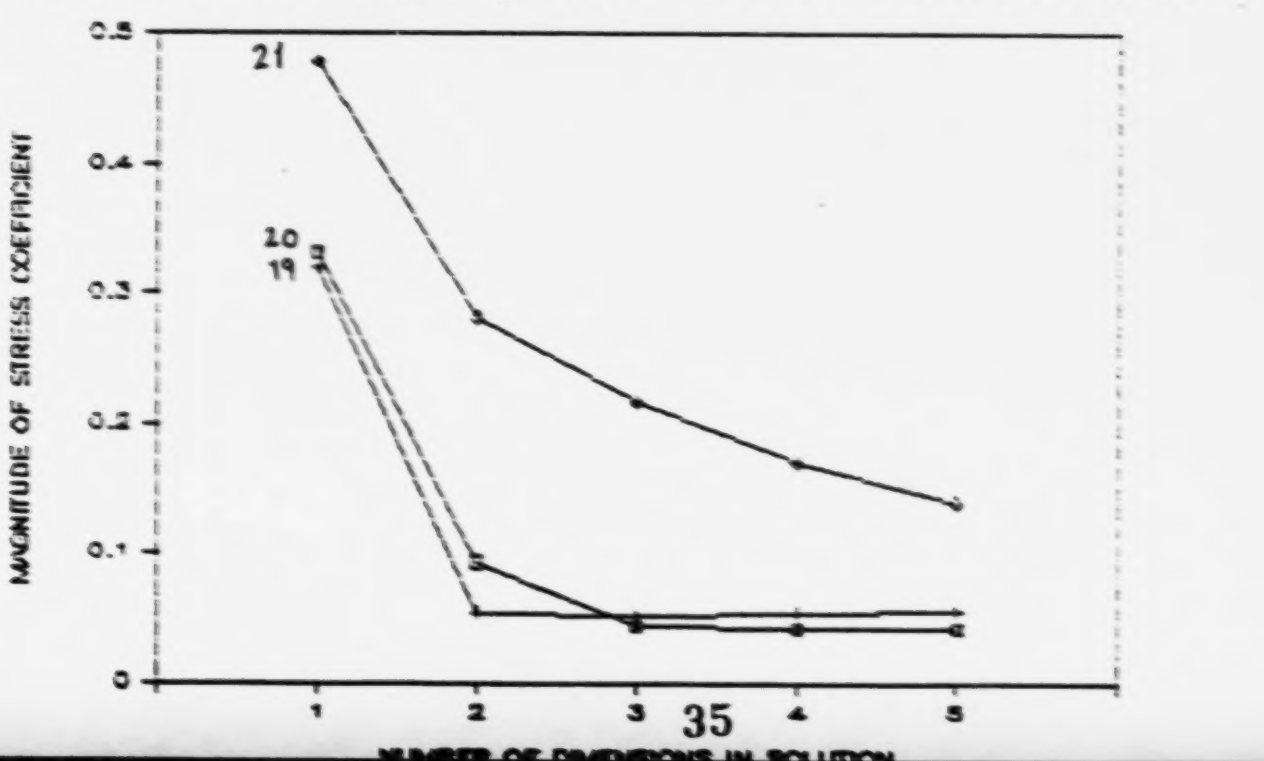


Figure 1.



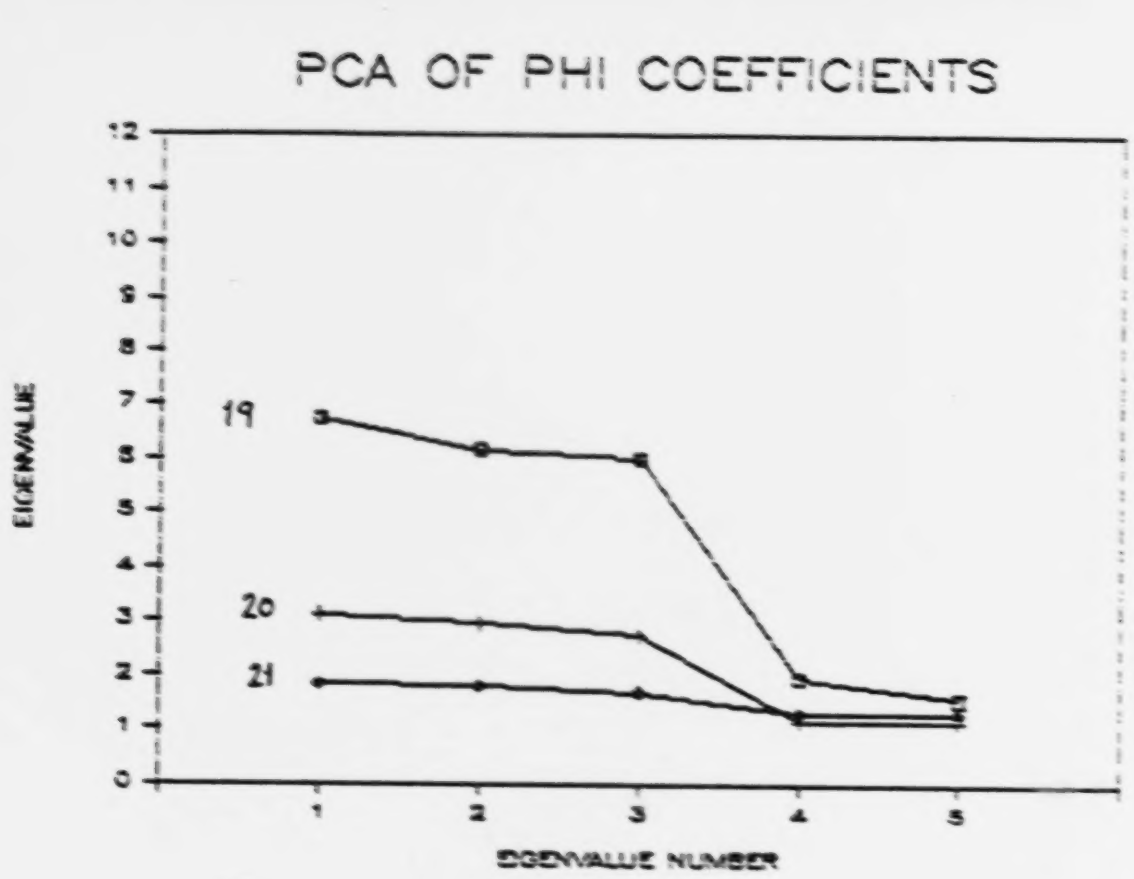
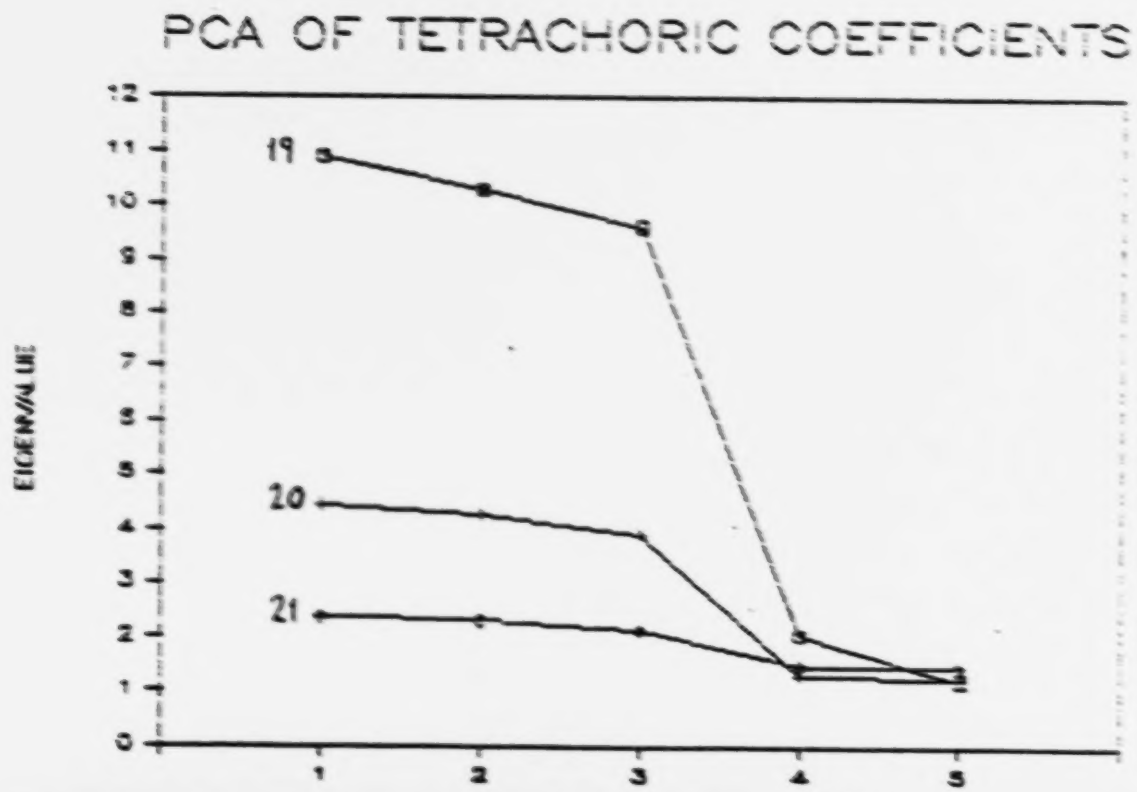

PCA OF PHI COEFFICIENTS
EIGENVALUE
EIGENVALUE NUMBER
19
20
21

Figure 3.


PCA OF TETRACHORIC COEFFICIENTS
EIGENVALUE
EIGENVALUE NUMBER
19
20
21

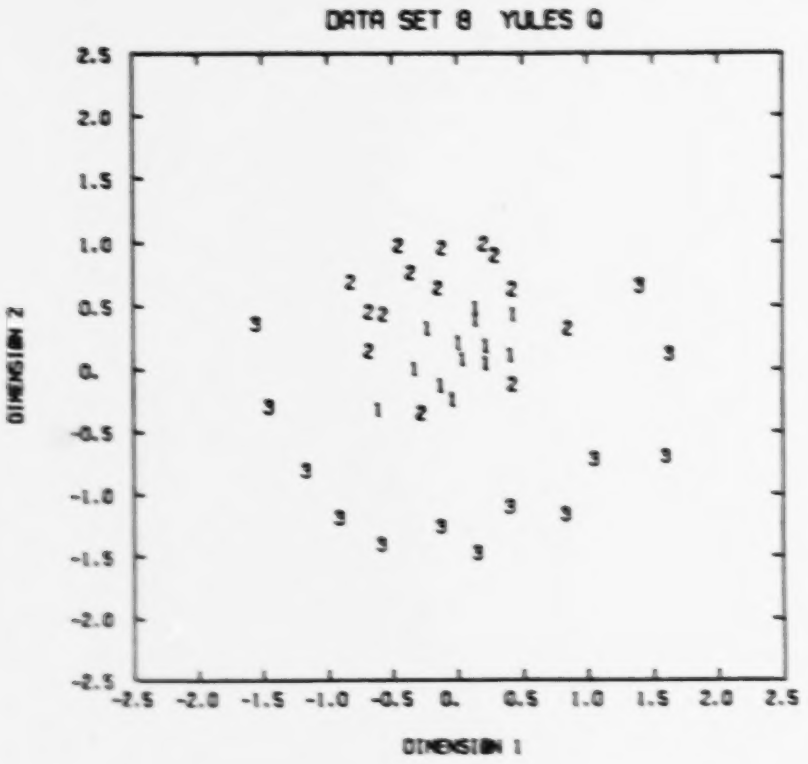


Figure 5.

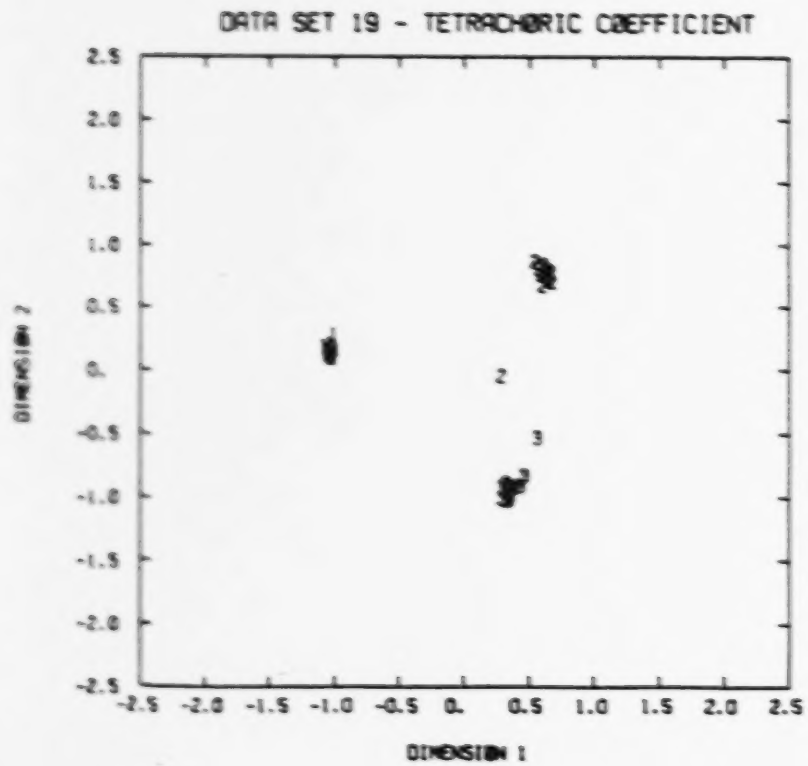


Figure 6.

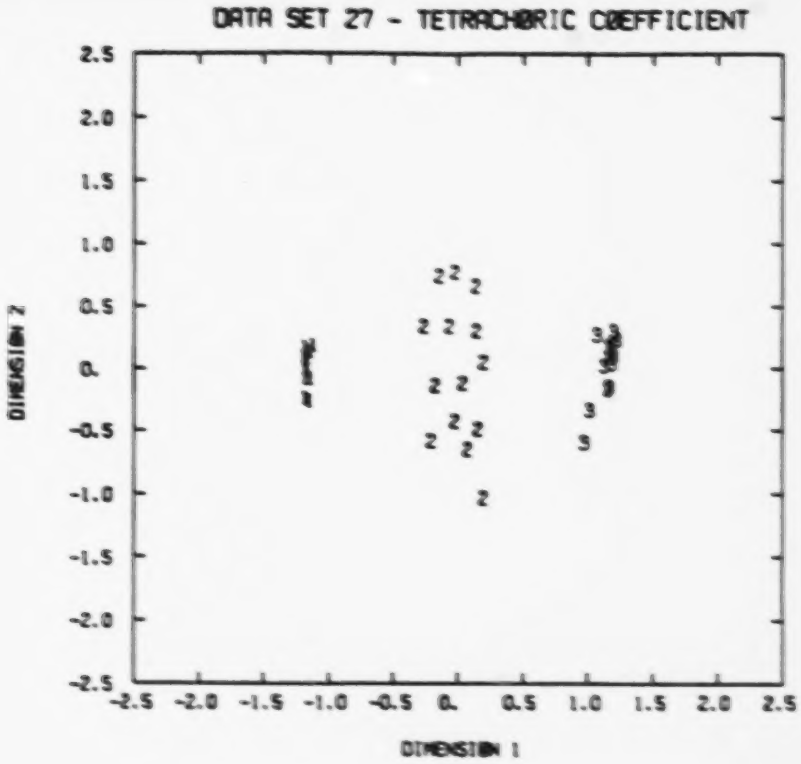


Figure 7.

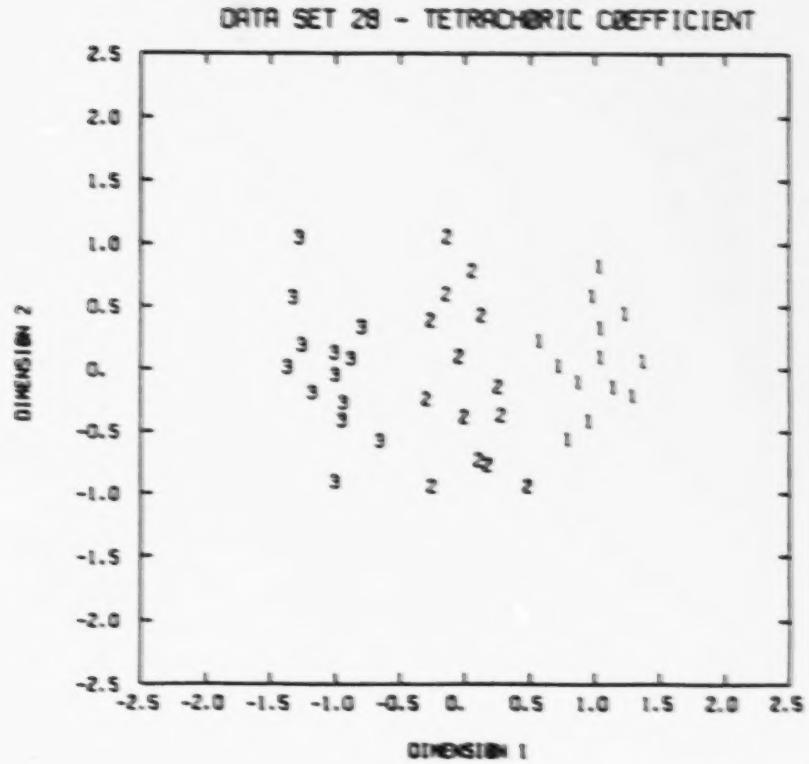


Figure 8.

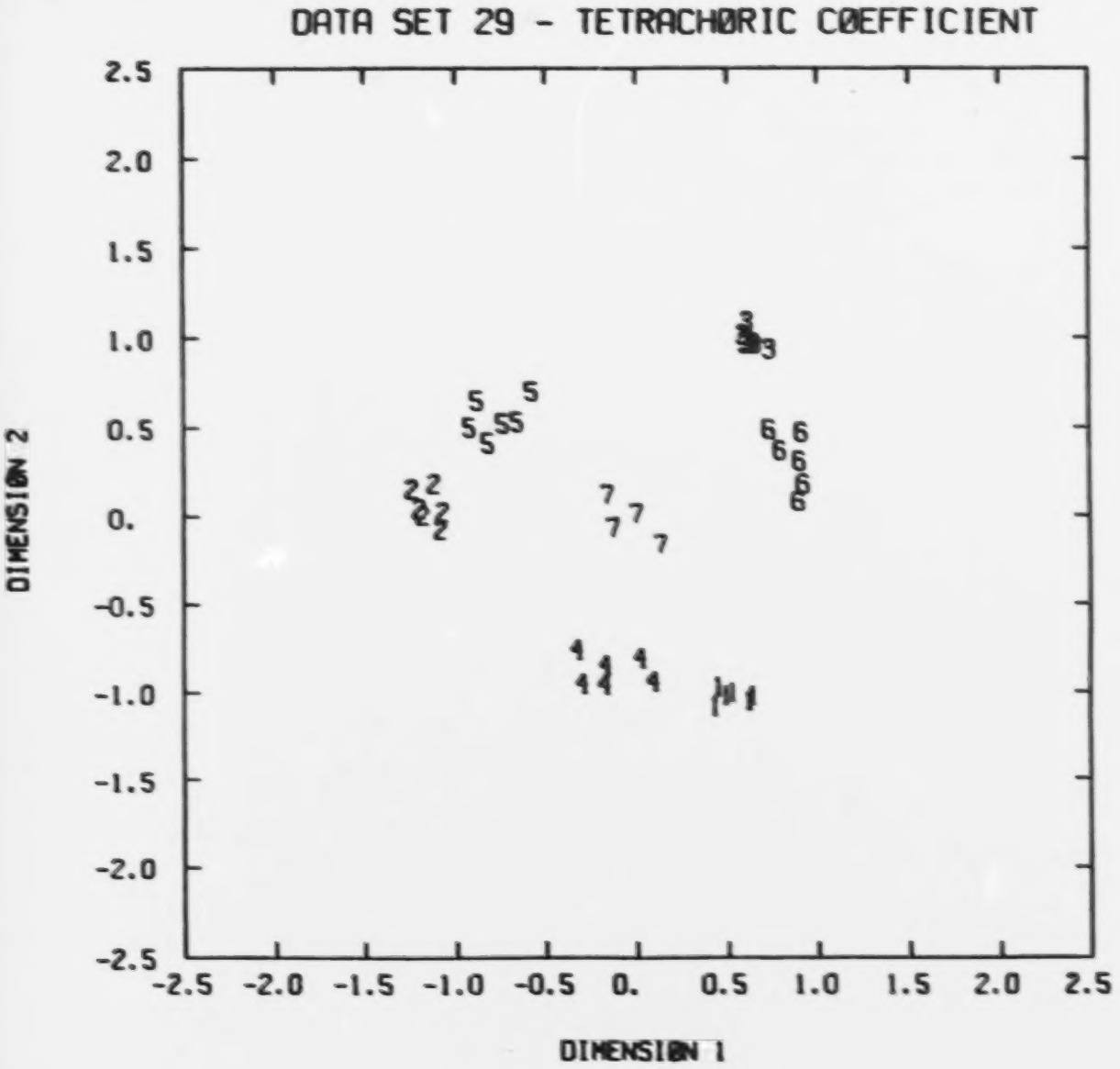

DATA SET 29 - TETRACHØRIC CØEFFICIENT
DIMENSIØN 2
DIMENSIØN 1
2.5
2.0
1.5
1.0
0.5
0.
-0.5
-1.0
-1.5
-2.0
-2.5
-2.5 -2.0 -1.5 -1.0 -0.5 0. 0.5 1.0 1.5 2.0 2.5

Figure 9.

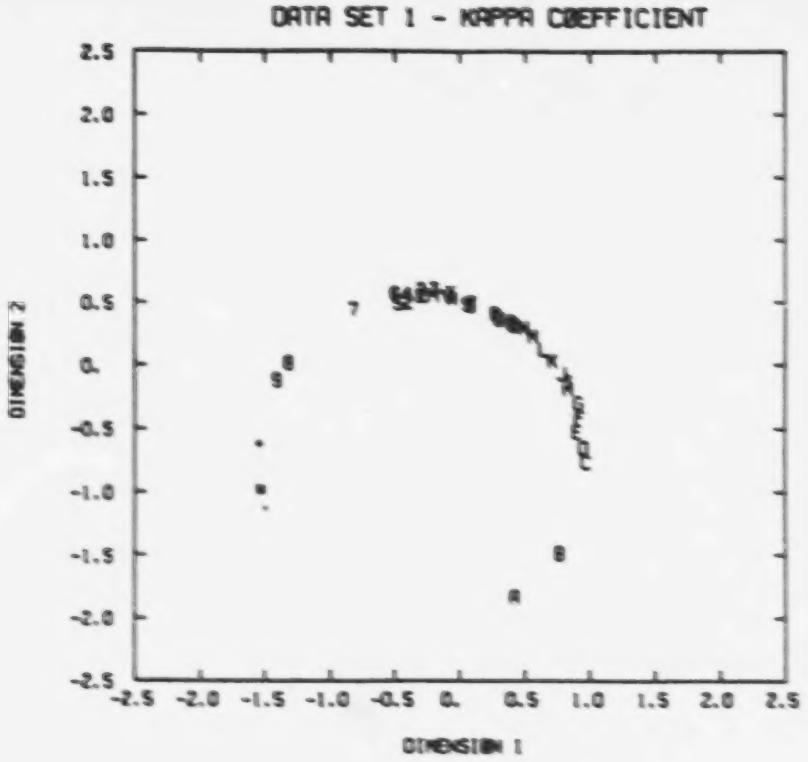


Figure 10.

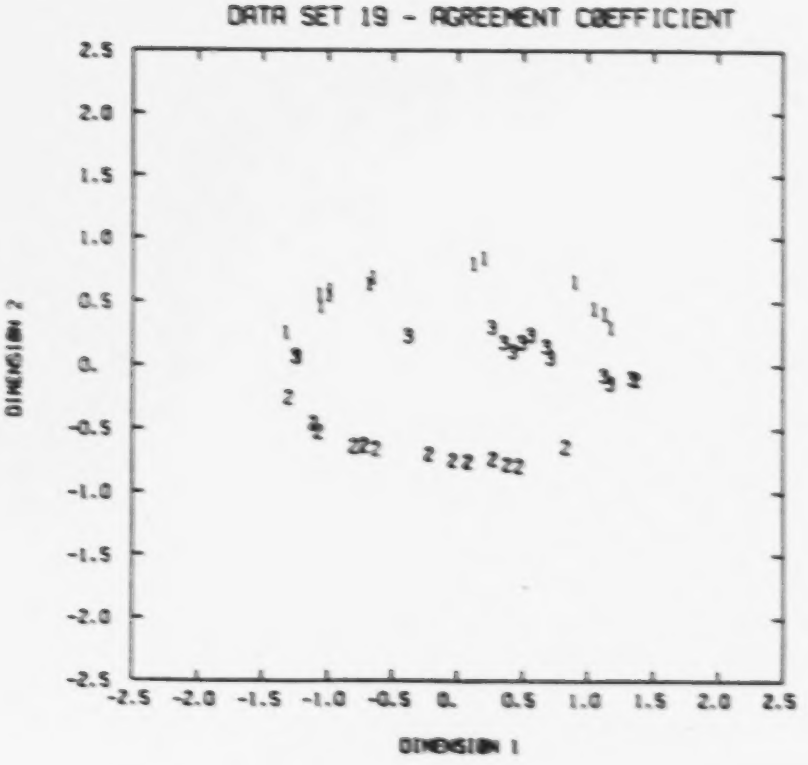


Figure 11.